Appendix: Handling Missing Clinical and Cost Data

Missing cost and clinical data were imputed using the Monte Carlo Markov Chain (MCMC) multiple imputation (MI) procedure, the last observation carried forward (LOCF) method, probabilistic imputation, or with negative values based on clinical expertise, as appropriate.

Specifically, for clinical variables in which fluctuations are normal and commonly observed clinically (i.e., ASDAS-CRP, BASDAI, BASFI, and HAQ), missing data was handled with MCMC imputation, which was considered the best strategy for representing this variability. For clinical variables observed to have greater stability, either the LOCF method or probabilistic imputation was used in lieu of MCMC imputation, as these methods provide better control over the variability of values imputed; LOCF was used if all patients had at least baseline data, while probabilistic imputation was used if baseline data were missing.

Following these rules, number of comorbid conditions and marital status were imputed by LOCF. Data on the satisfaction of ASAS imaging and ASAS clinical criteria were imputed probabilistically based on the distribution of non-missing values of those variables (i.e., the likelihood in the sample overall of satisfying these criteria). Data on the presence/absence of coxitis, IBD, uveitis, psoriasis/pustulosis were imputed with negative values based on clinical expertise, i.e., the rationale that these are relatively uncommon conditions, which may appear and disappear (meaning other strategies for dealing with missing data could result in excess positive values).

For missing cost data, separate assumptions were made for patients who attended the follow-up visit (partially incomplete data) and for patients who missed a follow-up visit (fully incomplete data). For patients who attended the visit, imputation of cost data was done only in cases where patients indicated resource use took place but did not provide sufficient details to allow for costing (i.e., missing or unintelligible responses to open-ended questions). This imputation was executed at the smallest resource component level and was done only for resource components for which either 1.5% or more of data were missing across all patients and visits, or for which less than 1.5% data of were missing but for which unit costs were high (e.g., surgery). For patients who attended the visit, no imputation was done for productivity loss costs (i.e., missing data was assumed to indicate no productivity loss). In cases where data were missing for professional occupation, values were imputed based on data from the next-closest visit (either LOCF or next observation carried back).

For patients who missed a visit and therefore had fully incomplete data, missing data were imputed at the following levels: anti-inflammatories; analgesics; corticoids; traditional DMARDs; biologics; medical acts; physician visits; all hospitalizations; productivity loss. The MCMC method was used to impute values for all resource components except biologics, for which the LOCF method was used, given the clinical observation that most patients remain on a stable dose of biologics following initiation.

For all cost data imputation, predictors entered into the MCMC model included health resource costs plus age, sex, marital status, ethnicity, disease duration, age at disease onset, ASDAS-CRP, BASDAI, BASFI, HAQ-AS, ASAS imaging and clinical criteria, peripheral arthritis at baseline, enthesitis at baseline, IBD (Crohn's disease or colitis), uveitis, psoriasis or pustulosis, coxitis grade (right and left) and comorbidity score (one point per self-reported comorbidity on a list of 23 comorbidities covering cardiovascular disease, gastrointestinal events, endocrine disorders, viral infections, tuberculosis, and Gougerot-Sjogren syndrome). For patients who missed the visit, the MCMC procedure considered non-missing data collected at previous visits. For patients who attended the visit, the MCMC procedure took into consideration non-missing data collected at previous visits and the most current visit. Data from 'future' visits were not taken into account when imputing data at a given time point.